



July 30, 2025

Sent via ECF

Hon. Ona T. Wang
Daniel Patrick Moynihan
United States Courthouse
500 Pearl Street
New York, NY 10007-1312

Orrick, Herrington & Sutcliffe LLP
The Orrick Building
405 Howard Street
San Francisco, CA 94105-2669
+1 415 773 5700
orrick.com

Annette L. Hurst
E ahurst@orrick.com
D +1 415 773 4585
F +1 415 773 5759

RE: *In re: OpenAI, Inc., Copyright Infringement Litigation*, No. 1:25-md-3143; this document relates to *The New York Times Co. v. Microsoft Corp., et al.* No. 1:23-cv-11195

Dear Judge Wang:

Microsoft respectfully requests a pre-motion conference concerning The New York Times Company's ("The Times") refusal to produce its ChatExplorer data stored in Sumo Logic—a 1.97 GB repository. This repository contains logs documenting The Times's use of ChatExplorer—an internally-developed Times AI product that uses the GPT models at issue in this case. Microsoft moves to compel The Times to produce all existing ChatExplorer data.¹

The Times's ChatExplorer data demonstrates it is using the accused GPT models for noninfringing, transformative, and pro-copyright purposes. As a commercial content producer, The Times's noninfringing and pro-copyright uses of the technology at issue are particularly relevant to Microsoft's substantial noninfringing uses and fair use defenses. The Times has failed to put forth any legitimate ground to refuse production, and indeed, previously agreed to produce at least a portion of the logs.

The data is relevant and there is little to no burden in producing less than 2 GB of user data in a case where Plaintiff has demanded inspection of vast multitudes of Defendants' user data. The parties conferred via multiple emails spanning months and via videoconference on July 24, but were unable to resolve this dispute. Ex. A (parties' correspondence).

The Data Microsoft Seeks. The Times developed a [REDACTED] called "ChatExplorer" for interacting with OpenAI's GPT models via [REDACTED] Ex. B ([REDACTED]) at 2; Ex. C ([REDACTED]). ChatExplorer is, in The Times's words, a [REDACTED] (Ex. D ([REDACTED]) at 4) that uses the GPT models at issue. The Times logs its ChatExplorer user interactions (*i.e.*, prompts from Times employees to OpenAI's GPT models, and the corresponding responses) in a central repository called Sumo

¹ Microsoft seeks what ChatExplorer data remains on Sumo Logic after The Times's deletions that are the subject of OpenAI's spoliation motion. *See* ECF 377.



Logic. *See, e.g.*, Ex. E (E. Mandel Dep. Tr.) highlighted portions). Thus, the ChatExplorer logs Microsoft seeks are a record of The Times’s use of OpenAI’s GPT models via [REDACTED].

The Times Previously Agreed to Produce ChatExplorer Documentation & A Subset of ChatExplorer Data. The Times’s document production, statements made during conferrals and to the Court make clear that The Times, through ChatExplorer, is using the accused GPT models [REDACTED]. *See, e.g.*, Ex. F ([REDACTED]) at 2; Ex. A; NYT Dkt. 429 (NYT Resp. to OpenAI’s Rule 72 Obj.) at 1-2, 5. After learning about the ChatExplorer data stored in Sumo Logic, Microsoft pressed on its targeted discovery requests to better understand what user data is stored in this platform. *See, e.g.*, MSFT RFP No. 106 (The Times’s use of its “internal ChatGPT equivalent”); *see also* MSFT RFP Nos. 229, 233 (similar). The Times repeatedly agreed to produce documents regarding its use of Defendants’ models. *See, e.g.*, NYT Dkt. 429 at 1-2, 5; Ex. G (A. Schreiber EM to A. Muttalib) at 1; Ex. H (NYT’s Resps. to MSFT’s RFP Nos. 229, 233). The ChatExplorer data logs are exactly that. The Times initially agreed to produce a subset of the ChatExplorer data and was willing to [REDACTED] Ex. A at 18. After this initial agreement, The Times did a 180-degree turnabout, reneging on its offer to produce and claiming the data was irrelevant. Ex. A at 1-2.

The Times’s Use of the GPT Models at Issue Supports Microsoft’s Substantial Noninfringing Uses Defense to The Times’s Contributory Infringement Claim. The Times is seeking to hold Microsoft secondarily liable for end-users’ use of Defendants’ generative AI tools. ECF 72 (Second Amended Compl.) ¶¶ 185-187. The Times hypothesizes that users generate infringing outputs, either through “regurgitated” answers from the GPT model or through prompts that involve use of its data as part of the grounding process associated with retrieval augmented generation. However, when a product is “capable of substantial noninfringing uses,” the provision “of such equipment to the general public does not constitute contributory infringement.” *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 456 (1984)); *Cartoon Network LP, LLLP v. CSC Holdings, Inc.*, 536 F.3d 121, 133 (2d Cir. 2008). Courts must “consider whether ... a significant number of” “different potential uses of the machine” “would be noninfringing.” *Sony*, 464 U.S. at 442. There is no better evidence of commercially significant noninfringing uses than how users such as The Times are in fact engaged in them—exactly what the data logs Microsoft seeks will show.

The Court’s prior orders assume that The Times would produce documents regarding its use of Defendants’ generative AI tools for journalism. *See, e.g.*, NYT Dkt. 429 at 1-2, 5 (asserting where “Defendants’ technology is used to aid Times reporting or its presentation of content, The Times has already agreed to produce documents.... The Times has agreed to produce responsive documents regarding its use of [Defendants’] tools for its journalism.”); ECF 82 (J. Stein Rule 72 Order) at 10-11 (relying on The Times’s assertions). Despite its concessions made in connection with those disputes, The Times now refuses to produce the logs recording its use of the GPT models at issue, which are “relevant evidence” to assess “the existence vel non of substantial noninfringing uses of defendants’ products.” ECF 82 at 10-11. The Times’s use of OpenAI’s GPT models via ChatExplorer “for research,” amongst other journalistic pursuits, would constitute a “substantial, predominant, and noninfringing use[]” and be strong evidence in support of a



complete defense to The Times’s contributory infringement allegations. *See Matthew Bender & Co. v. W. Publ. Co.*, 158 F.3d 693, 706–07 (2d. Cir. 1998) (holding no contributory liability on these grounds); *see also MGM Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, 933 (2005). Accordingly, the ChatExplorer data reflecting such uses is highly relevant and discoverable.

The Times’s Use of the GPT Models at Issue is Also Relevant to Fair Use. The ChatExplorer logs will also provide evidence that the accused products are used for transformative purposes. The first factor concerns whether an accused product “makes some contribution of new intellectual value and thereby foster[s] the advancement of the arts and sciences[.]” *Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 923 (2d Cir. 1994). Courts determine that accused products are transformative when this constitutional objective has been sufficiently met. *See Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994) (“transformative” uses include those that “further[.]” “the goal of copyright, to promote science and the arts”). We expect The Times, via ChatExplorer, is using the accused GPT models “for purposes such as criticism, comment, news reporting,” and “research.” 17 U.S.C. § 107. Indeed, The Times has admitted to using “OpenAI’s Large Language models to help with brainstorming and editing, and to summarize Times-owned materials[.]” Ex. I (NYT’s Resps. to OpenAI’s Reqs. for Admis. Nos. 97-120). Further, The Times’s [REDACTED]

[REDACTED] Ex. B at 2. These are all uses that fulfill the “basic constitutional objective of copyright,” and demonstrate how the accused technology is transformative. *See, e.g., Google LLC v. Oracle Am., Inc.*, 141 S. Ct. 1183, 1203-04 (2021) (Google’s use of copyrighted works “to create new products” was “consistent with that creative ‘progress’” and thus were transformative).

There is No Burden for The Times to Produce 1.97 GB of Non-Privileged Data. The Times has not articulated any meaningful burden to produce what amounts to less than a thumb drive of data: 1.97 GB, consisting of approximately 82,150 requests logged. Ex. J (NYT’s Resps. to OpenAI’s Preservation Questions) at 1. The Times hypothesizes its data *may* contain privileged information, but it performed no investigation to substantiate its assertion. Further, it’s unlikely privileged ChatExplorer data exists because The Times confirmed its legal department does not use ChatExplorer. *Id.*, at 3.

The Times atmospherically asserts reporters’ privilege on the sole ground that its journalists use ChatExplorer. Yet, The Times did not investigate whether the data logs identify confidential sources, which is unlikely given [REDACTED]

[REDACTED] Ex. B at 2; Ex. C ([REDACTED]); Ex. K ([REDACTED]) at 5.

While it seems doubtful a journalist would input a confidential source into a third-party model, The Times can withhold such data and log these incidents. Any data produced would be governed by the Protective Order; and, in all events it is hard to see how this “qualified privilege” designed to “protect[] journalists’ sources[,]” would apply to Plaintiff’s use of the accused tools. *N.Y. Times Co. v. Gonzales*, 459 F.3d 160, 180 (2d Cir. 2006) (quoting *Baker v. F & F Inv.*, 470 F.2d 778, 781-82 (2d Cir. 1972)). The Times should be compelled to produce the ChatExplorer data in full.



Respectfully submitted,

/s/ Annette L. Hurst

Annette L. Hurst